BITCOIN PRICE FORECASTING TIME SERIES APPROACH

June 2023

Authors: Hera K.

1. INTRODUCTION

This project aims to forecast future bitcoin price based on historical bitcoin price data by time series approach. In time series analysis, time is a crucial variable of the data, because it shows us how the data adjusts over the course of the data points as well as the result.

Time series is a specific way of analyzing a sequence. It can show likely changes in the data, like seasonality or cyclic behavior, which provides a better understanding of data variables and helps forecast better.

2. THE BUSINESS PROBLEM

Bitcoin emerged as a virtual currency to challenge the existing centralized system after 2008 financial crisis. Bitcoin the first and most well-known cryptocurrency was created by a Japanese American man living in California, in 2009 by an individual or group of individuals using the pseudonym Satoshi Nakamoto.

Bitcoin is created, distributed, traded, and stored using a decentralized ledger system known as a blockchain. Bitcoin and other cryptocurrencies are not currently considered real money by the federal reserve or U.S. banks. However, it's managed by blockchain technology and verified by all users on the network.

The main difference between the traditional currency and bitcoins are that traditional currency is a centralized system and bitcoins are decentralized one and peer-peer systems without any intermediary. No single person or group has control—instead, all users collectively retain control.

In a traditional banking system, for a national transaction takes 2-3 working days, and the transaction fees will be high. In the case of international transactions, the transaction fee will be much higher, and it will take 15 days to complete the transaction. In a Cryptocurrency system like bitcoins, there is no transaction fee for both national transaction and globally. The transaction will also take place in seconds or within 24 hours, as a bitcoin system function 24/ 7. Because this blockchain system is a distributed, immutable, and decentralized ledger at its core that consists of a chain of blocks and each block contains a set of data. The blocks are linked together using cryptographic techniques and form a chronological chain of information. Decentralized blockchains are immutable, which means that the data entered is irreversible. For Bitcoin, transactions are permanently recorded and viewable to anyone.

A. Work System of Bitcoin

i. In short, a centralized digital network is a network in which collective data is processed and stored by a single node. Conversely, a decentralized digital network is a network in which multiple nodes store and process data. A store of value is an asset, currency, or commodity that maintains its value over a long period. An item would be considered a store of value if its value is either stable or increases over time but doesn't depreciate. Bitcoin is proving to be a legitimate store of value is its scarcity.

ii. Cash is issued by a government, cryptocurrency is not. Crypto isn't controlled by an individual, institution, or any other authority.

B. History of Bitcoin

i. Highlighted outstanding events chrolonogically.

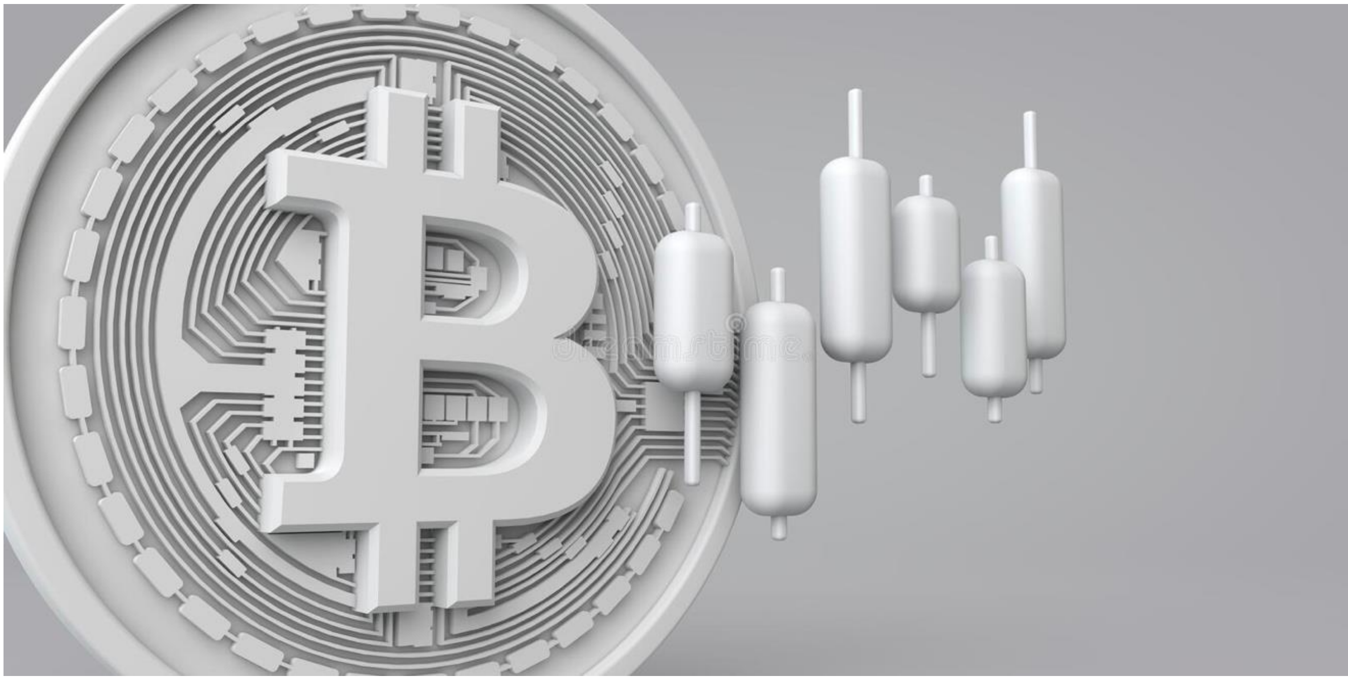
- *2008 – Invented by Satoshi Nakamoto*
- *2009 – The bitcoin began its use a open-source*
- *2011 – The largest decrease occurred as a percentage of the coin's value*
- *2013 – BTC is $ 1 in globe market*
- *2017 – A rise in prices, bitcoins were worth ±$998*
- *2021 – El Salvador was the first country who adopted BTC as legal tender*

- *2023 – Tesla sold 75 % of its BTC holdings changed the market*

C. Difference Between Crypto & Currency

i. The most important difference is centralized- decentralized system as mentioned above particularly. Difference of their storage determines the value of asset.
ii. Bitcoin is limitless and centered globally, cash is dependent on government's policy or other lots of factors.
iii. Cash system has brought about significant obstacles resulting from pandemic crisis, then global economic downturn, economy has been almost over.

To come up with a long-lasting solution, the world needs innovative and robust perspective for global system. Increasing demand worldwide makes Bitcoin much more attractive, though it looks like untrusty and jeopardous venture for now. Bitcoin prices are tried to be predicted by investors or majority. Bitcoin is aspiring critical up-and-coming for future as it ensures an analytical, favorable, strategic system emerging in digitalized world. To forecast bitcoin future prices will guide for more entrepreneurs in finance world, it will be enabled to invest wisely.



3. THE DATA

I have bitcoin dataset from Bitcoin yahoo finance source, the dataset consists of date, opening, closing prices of bitcoin, high and low price as well, adjusted closing price, volume from September 2014 to May 2023.

I have four targets: open, close, low and high prices. Columns in my dataframe are explained as below. Bitcoin price operates 24 hours a day nonstop. All entries emphasizes daily prices for this dataset.

Date: Index in our time series that specifies the date associated with the price. (USD)

Open Price: The first price of BTC was purchased on the trading day (USD)

Close Price: The last price of BTC was purchased at the end of trading day (USD)

High: The maximum price of BTC was purchased on trading day (USD)

Low: The minimum price of BTC was purchased on the trading day (USD)

Adjusted Closing Price: Stock exchanges witness buying and selling of millions of shares every minute. When the exchanges close, the last trading price of the stock is recorded as the closing price of the share. Data is adjusted using appropriate split and dividend multipliers, adhering to Center for Research in Security Prices (CRSP) standards (USD)

Volume: The sum of actual trades made during the trading day (USD)

- All Time High $68,789.63 USD. in 2021-11-10
- Max Volume 350,967,941.479 USD in 2021-02-26

A. Data Preparation

i. I call my csv file with parse_date method, it changes the given string from a datetime to the desired format into the dataframe. Then I implement EDA (Exploratory Data analysis)

ii. I check duplicated columns, null/ missing info, general dataframe and columns information.

iii. I visualize significant points to explore the data more interpretably for modelling.

iv. I check my data stationary thanks to the p-value by Dickey Fuller test result, then I implement ACF, PACF plots before modelling and to determine which terms are needed, what kind of model I would use. (The details and graphs are available in method and model parts).

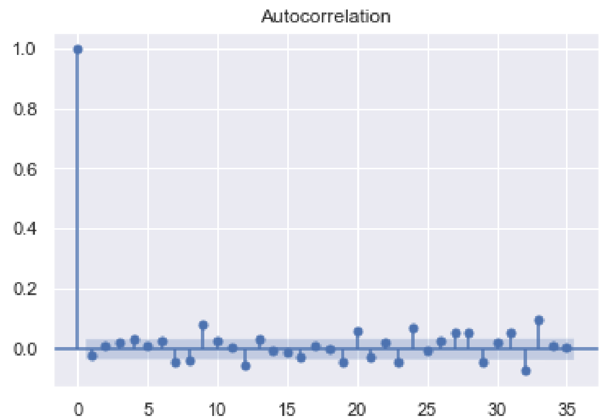
4. METHODS & MODELS

Time series method and model used for this analysis. I would like to mention about the methods and models I use for this analysis before dig into my model details.

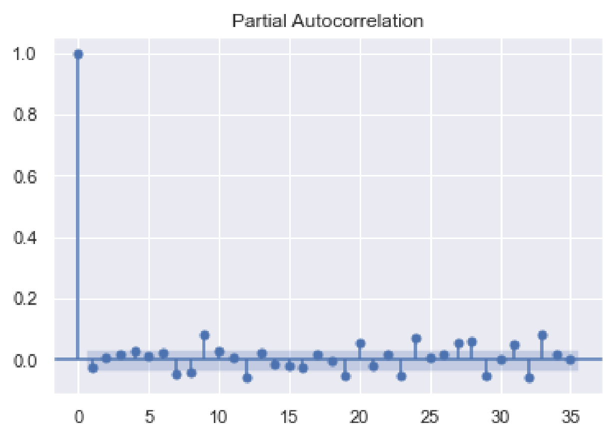
I focus on daily price at first glance, because daily open and close price causes to get profit. I don’t need to resample it because it is already my daily closing price is provided in my original dataset. Because of instability in my dataset. I can apparently see trend, seasonality on my visualization.

I check whether my data is stationary via ACF, PACF graphs and Dickey Fuller Test.

ACF is an Autocorrelation Function is that the correlation between the observation at the current time spot and the observations at previous time spots. Its plot provides answers about if the observed time series white noise or random, makes us to decide if I need to use MA term (q) in my model.



PACF is a Partial Autocorrelation Function is a statistically technique to assess the direct link between an observation at a given lag and its preceding delays. That means it subtracts the impact of the further delays between two-time series to assess the correlation between their values at a certain lag. It helps to decide for AR term (p).



Furthermore, it displays a **blue area** in the ACF and PACF plots, which depicts the 95% confidence interval and is in indicator for the **significance threshold**. That means, anything within the blue area is statistically close to zero and anything outside the blue area is statistically non-zero.

	AR(p)	MA(q)	ARMA(p, q)
ACF	Tails off (Geometric decay)	Significant at lag q / Cuts off after lag q	Tails off (Geometric decay)
PACF	Significant at each lag p / Cuts off after lag p	Tails off (Geometric decay)	Tails off (Geometric decay)

The augmented Dickey-Fuller test is an extension of the standard Dickey-Fuller test, which also checks for both stationarity and non-stationarity in the time series. My closing price is obviously non-stationary, to make stationary for the best forecasting I difference it. I used differencing method thanks to the ARIMA and SARIMAX Integrated term. Because of the non-stationary I need to specify AR and MA terms for my model.

I use ARIMA (Autoregressive Integrated Moving Average model and SARIMAX (Seasonal Autoregressive Integrated Moving Average) method for daily closing price.

ARMA model

I use the same method for 'Monthly Closing Price'. I resample my daily closing price as ' Monthly'. It is non-stationary because of outlier or booming BTC price in 2021 abruptly.

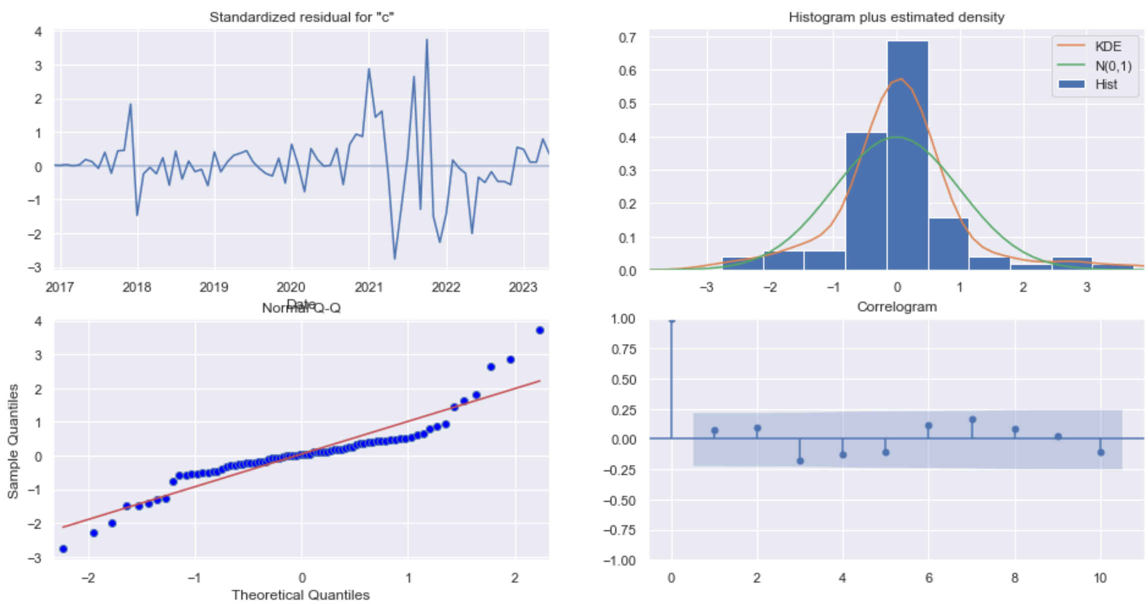
```
]: # Resample closing price Monthly

close_mts = df['close'].resample('M').mean()
close_mts
```

```
]: Date
2014-09-30      407.182428
2014-10-31      364.148873
2014-11-30      366.099799
2014-12-31      341.267871
2015-01-31      248.782547
                   ...
2023-01-31    20250.717490
2023-02-28    23304.539202
2023-03-31    25116.900895
2023-04-30    28857.574544
2023-05-31    28824.324609
Freq: M, Name: close, Length: 105, dtype: float64
```

I decompose my monthly closing time series into distinct components such as trend, seasonality, and noise.

The below diagnostic plot is not perfect; however, my model diagnostics suggest that the model residuals are almost close to normally distributed.

```
from pylab import rcParams
rcParams['figure.figsize'] = 18, 8
decomposition = seasonal_decompose(close_mts, model='additive')
fig = decomposition.plot()

#Save the figure
plt.savefig("Decomposition Monthly BTC.png", transparent = True, dpi=150)
plt.show()
```



SARIMAX Model

```
]: # Let's find the best AIC score and the mosrt optimal parameter combination among them for the
for param in pdq:
    for param_seasonal in seasonal_pdq:
        try:
            mod = SARIMAX(close_mts,
                          order=param,
                          seasonal_order=param_seasonal,
                          enforce_stationarity=False,
                          enforce_invertibility=False)
            results = mod.fit()
            print('ARIMA{}x{}12 - AIC:{}'.format(param, param_seasonal, results.aic))
        except:
            continue
```

```
ARIMA(0, 0, 0)x(0, 0, 0, 12)12 - AIC:2366.3749005646578
ARIMA(0, 0, 0)x(0, 0, 1, 12)12 - AIC:2101.1518679355577
ARIMA(0, 0, 0)x(0, 1, 0, 12)12 - AIC:2054.257402019919
ARIMA(0, 0, 0)x(0, 1, 1, 12)12 - AIC:1794.776092510484
ARIMA(0, 0, 0)x(1, 0, 0, 12)12 - AIC:2067.3070124300016
ARIMA(0, 0, 0)x(1, 0, 1, 12)12 - AIC:2044.4175901784024
ARIMA(0, 0, 0)x(1, 1, 0, 12)12 - AIC:1811.5238271387593
ARIMA(0, 0, 0)x(1, 1, 1, 12)12 - AIC:1793.1385592097708
ARIMA(0, 0, 1)x(0, 0, 0, 12)12 - AIC:2269.6502099982126
ARIMA(0, 0, 1)x(0, 0, 1, 12)12 - AIC:1999.7629988833535
ARIMA(0, 0, 1)x(0, 1, 0, 12)12 - AIC:1958.8213261003343
ARIMA(0, 0, 1)x(0, 1, 1, 12)12 - AIC:1703.3252856342697
ARIMA(0, 0, 1)x(1, 0, 0, 12)12 - AIC:2040.3454698244498
ARIMA(0, 0, 1)x(1, 0, 1, 12)12 - AIC:1910.4780512847124
ARIMA(0, 0, 1)x(1, 1, 0, 12)12 - AIC:1746.5168754481208
ARIMA(0, 0, 1)x(1, 1, 1, 12)12 - AIC:1705.322250943111
ARIMA(0, 1, 0)x(0, 0, 0, 12)12 - AIC:1991.079991261026
ARIMA(0, 1, 0)x(0, 0, 1, 12)12 - AIC:1776.2838513551467
ARIMA(0, 1, 0)x(0, 1, 0, 12)12 - AIC:1828.9518134019052
ARIMA(0, 1, 0)x(0, 1, 1, 12)12 - AIC:1548.7288070841078
```

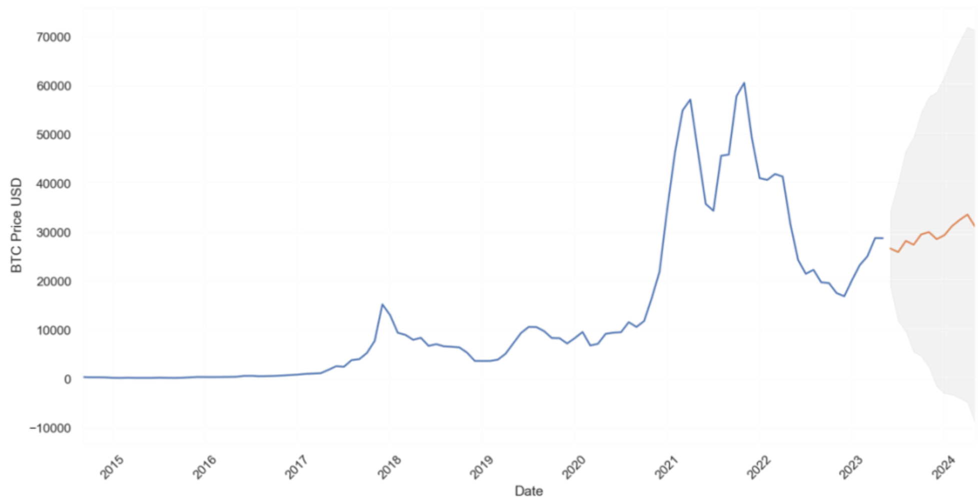
```
# The lowest AIC, ARIMA(0, 1, 1)x(1, 1, 1, 12)12 - AIC:1510.1090594876027

mod = SARIMAX(close_mts,
              order=(0, 1, 1),
              seasonal_order=(1, 1, 1, 12),
              enforce_stationarity=False,
              enforce_invertibility=False)

results = mod.fit()
print(results.summary())
```

```
                                      SARIMAX Results
==========================================================================================
Dep. Variable:                                close   No. Observations:                  105
Model:             SARIMAX(0, 1, 1)x(1, 1, 1, 12)   Log Likelihood                -751.055
Date:                              Fri, 26 May 2023   AIC                           1510.109
Time:                                      13:23:13   BIC                           1519.536
Sample:                                  09-30-2014   HQIC                          1513.883
                                       - 05-31-2023
Covariance Type:                                opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ma.L1          0.5539      0.058      9.498      0.000       0.440       0.668
ar.S.L12      -0.0045      0.149     -0.030      0.976      -0.297       0.288
ma.S.L12      -1.3041      0.072    -18.037      0.000      -1.446      -1.162
sigma2      9.198e+06   1.14e-08   8.05e+14      0.000      9.2e+06     9.2e+06
===================================================================================
Ljung-Box (L1) (Q):                   0.52   Jarque-Bera (JB):                46.25
Prob(Q):                              0.47   Prob(JB):                         0.00
Heteroskedasticity (H):               6.32   Skew:                             0.64
Prob(H) (two-sided):                  0.00   Kurtosis:                         6.55
===================================================================================
```

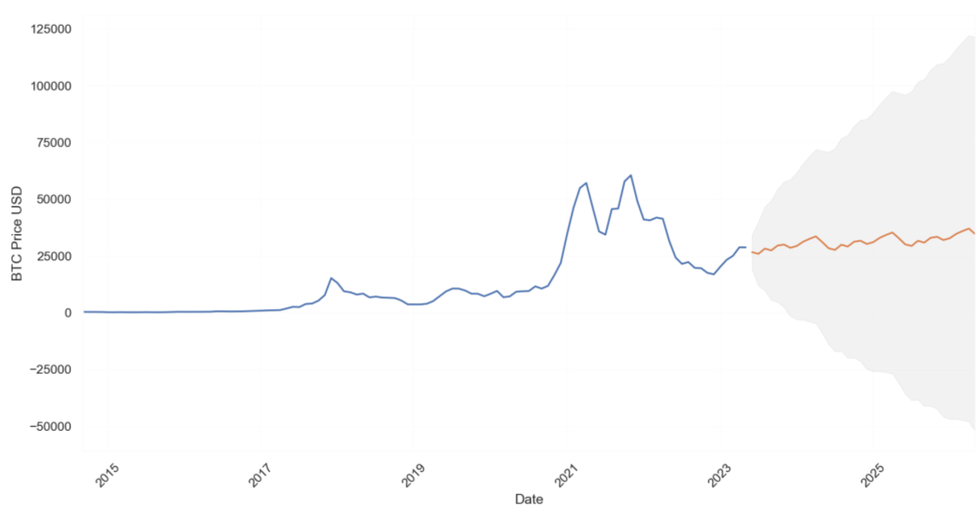
RMSE : 3230.9 (Expected Root Mean Squared Error)
AIC : 1510. 1 (Akaike Information Criteria)

RMSE : 4333.3 (Forecast Root Mean Squared Error)

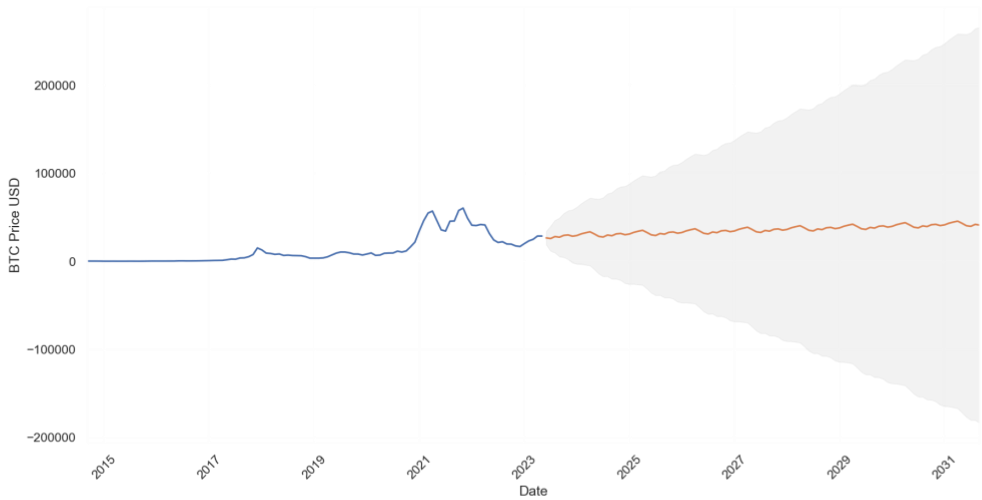
Forecast 1 Yearly BTC Price (USD)



Forecast 3 Yearly BTC Price (USD)



Forecast 8 Yearly BTC Price (USD)



5. EVALUATION

I report both the RMSE and AIC……… project aims to forecast future bitcoin price based on historical bitcoin price data by time series approach. In time series analysis, time is a crucial variable of the data, because it shows us how the data adjusts over the course of the data points as well as the final result.

6. FURTHERMORE

There are other factors to specify the prediction of BTC price. This analysis is only based on change the price over time as mentioned above.

Other reasons mentioned below would be concerned as well.

- Price swings in bitcoin are mainly driven by its own halving cycles as well as macroeconomic events.
- Global effects
- Inflation, interest rates etc.
- Social platform in business
- I analyze only 10 years background. BTC is just emerging, developing, if there is more year's dataset, it causes the more clear prediction result.

7. REPOSITORY STRUCTURE

```
[Data]
 │      ├── BTC-USD.csv (https://finance.yahoo.com/quote/BTC-USD/history/)
 ├── [Images]
 ├── [Model]
 ├── .gitignore
 ├── Bitcoin_Price_Prepwork.ipynb
 ├── LICENSE
 ├── Presentation.pdf
 ├── README.md
 ├── environment.yml
 ├──
 └──
```